DOCUMENT RESUME

ED 167 384 SE 026 761

AUTHOR Helgeson, Stanley L., Ed.; Blosser, Patricia E., Ed.
TITLE Investigations in Science Education, Vol. 4, No. 2. [Expanded Abstracts and Critical Analyses of Recent Research.]
INSTITUTION Ohio State Univ., Columbus. Center for Science and Mathematics Education.
PUB DATE 78
NOTE 68p.
AVAILABLE FROM Information Reference Center (ERIC/IRC), The Ohio State University, 1200 Chambers Rd., 3rd Floor, Columbus, Ohio 43212 (Subscription $6.00, $1.75 single copy)

EDRS PRICE MF-$0.83 HC-$3.50 Plus Postage.
DESCRIPTORS *Cognitive Development; *Educational Research; *Instruction; Research Methodology; Science Education; *Teacher Education
IDENTIFIERS *Piaget; Research Reports

ABSTRACT

This issue of "Investigations in Science Education," provides analytical abstracts, prepared by science educators, of research reports in the areas of instruction, Piagetian studies, and teacher education. Each abstract includes bibliographical data, research design and procedure, purpose, research rationale, and an abstractor's analysis of the research. (BB)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

- ED167384

INVESTIGATIONS IN SCIENCE EDUCATION

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATIC ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

Editor

Stanley L. Helgeson
The Ohio State University

Associate Editor

Patricia E. Blosser
The Ohio State University

Advisory Board

David P. Butts (1978)
University of Georgia

Kenneth G. Jacknicke (1978)
University of Alberta

Donald E. Riechard (1979)
Emory University

Ronald D. Anderson (1981)
University of Colorado

Frances Lawrenz (1980)
Minneapolis, Minnesota

Joe C. Long (1981)
University of Georgia

National Association for Research in Science Teaching

ERIC Clearinghouse for Science, Mathematics,
and Environmental Education

Published Quarterly by

The Center for Science and Mathematics Education
College of Education
The Ohio State University
1945 North High Street
Columbus, Ohio 43210

Subscription Price: $6.00 per year. Single Copy Price: $1.75.
Add 25¢ for Canadian mailings and 50¢ for foreign mailings.

SE 026 761

Volume 4, Number 2, 1978

NOTES from the Editor iii

ERRATA iv

INSTRUCTION 1

DeTure, Linda R. and John J. Koran, Jr. "Using Filmed Models to Prompt Children's Laboratory Activities." Journal of Research in Science Teaching, 12: 449-452, 1975.
Abstracted by WILLIAM G. HOLLIDAY AND JUDITH A. THREADGILL 3

Hunt, John D. and Bill Lamkin. "An Analysis of Reinforcement and Feedback within an Auto-Tutorial Plant Taxonomy Module." Journal of Research in Science Teaching, 12(2): 121-125, 1975.
Abstracted by GENE F. CRAVEN 9

Kren, Sandra R. and John P. Huntsberger. "Should Science Be Used to Teach Mathematical Skills?" Journal of Research in Science Teaching, 14(6): 557-561, 1977.
Abstracted by GEORGE W. BRIGHT 17

Valeriote, I. M. "A Self-Paced Laboratory Course in First-Year General Chemistry." Journal of Chemical Education, 53(2): 106-108, 1976.
Abstracted by J. DUDLEY HERRON 21

Wideen, Marvin F. and Richard L. Butt. "Student and Instructor Directed Approaches to Elementary Science Education: A Comparative Analysis." Journal of Research in Science Teaching, 12(2): 95-100, 1975.
Abstracted by JOHN D. BALLING. 29

PIAGETIAN STUDIES 37

Lawson, A.E. and J.W. Renner. "A Quantitative Analysis of Responses to Piagetian Tasks and Its Implications for Curriculum." Science Education, 58(4): 454-559, 1974.
Abstracted by FRED W. FINLEY 39

Schafer, L.E. and J.L. Byers. "The Effectiveness of Cue Fading in Teaching Kindergarten Children to Serial Order." Journal of Research in Science Teaching, 12(3): 281-292, 1975
Abstracted by WILLIS J. HORAK. 46

TEACHER EDUCATION 53

Chrouser, William H. "Outdoor vs. Indoor Laboratory Techniques in Teaching Biology to Prospective Elementary Teachers." Journal of Research in Science Teaching, 12(1): 41-48, 1975.
Abstracted by WILLIAM R. BROWN 55

Greenwood, Gordon E. and Richard R. Renner. "Student Ratings of College Teaching: Their Use in Administrative Decision Making." Science Education, 59(4): 493-498, 1975.
Abstracted by HERBERT A. SMITH 62

NOTES . . .

from the Editor

With this issue of INVESTIGATIONS IN SCIENCE EDUCATION, the present Editor will conclude his responsiblities in this capacity. Beginning with Volume 4, Number 3, Patricia E. Blosser will assume responsibility as Editor and Robert L. Steiner will serve as Associate Editor.

I would like to take this opportunity to thank the National Association for Research in Science Teaching, the abstractors who have devoted their time and efforts to this journal, and to my colleagues, all of whom have helped in bringing this publication into being. As we all know, such an undertaking has its hazards, errors, and circumstances of chaos, yet there are the rewards of assisting dialogue within the profession and in providing a forum aimed at improving the reporting of research, and ultimately the research itself, in science education.

Best wishes to Pat and Bob as they take over responsibility for this journal and its continued improvement.

Stanley L. Helgeson
Editor

Patricia E. Blosser
Associate Editor

ERRATA

We wish to call your attention to an error, on page 64, in Volume 3, No. 1, 1977. The second line from the bottom of page 64 should read as follows:

concept attainment) would have been significant, merely because of . . .

(When this was printed, the adjective insignificant was substituted for the correct term, significant.)

INSTRUCTION

DeTure, Linda R. and John J. Koran, Jr. "Using Filmed Models to Prompt Children's Laboratory Activities." Journal of Research in Science Teaching, 12:449-452, 1975.

Descriptors--Educational Research; *Elementary Education; Elementary School Science; *Films; *Instruction; *Observational Learning; *Peer Teaching; Science Activities; Science Education

Expanded Abstract and Analysis Prepared Especially for I.S.E. by William G. Holliday and Judith A. Threadgill, University of Calgary.

Purpose

The major objective of this study was to evaluate observation of peer group behaviors as a method of teaching procedural techniques for performing a science experiment. A second objective was to explore the feasibility of conducting modeling research of this type in a school setting.

Rationale

Students exposed to an activity-oriented science program often fail to understand the procedure required to execute the lessons, and are thereby severely limited in their chances of reaching the original curricular objectives. Providing students with an opportunity to observe someone else perform a procedure in a clear, direct, and unambiguous fashion can be an effective method of overcoming such procedural difficulties. Observational learning can be an even more powerful method when the learner observes a prestigious person executing a set of behaviors. Therefore, students who observe a popular group of peers perform activity procedures are more likely to learn those techniques, execute the science experiment, and acquire associated process skills than students who merely listen to a teacher verbally describe the same techniques.

Research Design and Procedure

The 22 fourth-grade students participating in this study were from a school serving a poor community. These students (half of which were

black; half, white) were identified as coming from low socioeconomic families. Prior to the study, it was observed that these students had difficulty in following the teacher's verbal directions on correct experimental procedures.

Four students constituting the modeling group were selected on the basis of peers' responses to two questionnaires: "How I Feel About Others In My Class" and "The People In My Class." These instruments were developed by Fox, Luski and Schmuck (1966) and were designed to identify the best liked and most powerful students in a classroom as judged by the group. In this instance, the four students in the modeling group all received substantially higher positive ratings from the questionnaires than did the other students.

The four students selected as models were instructed by their teacher in those experimental techniques required to complete the objectives described in a Science—A Process Approach lesson. The model group then physically performed the experiment and verbally explained their performance before a video-tape camera operated by one of the investigators. It was hypothesized that the treatment students viewing the video-tape would enjoy seeing their popular friends on "television" and learn the presented technique behaviors.

Verbal instructions concerning the activity procedures and the demonstration of safety rules were described by the teacher to all students during a 15-minute introduction. The 18 students were then randomly assigned to either a control group or a treatment group, each with nine students. Three subgroups of students were identified in each major group and three sets of specific tasks related to the lesson procedure were assigned to members of each subgroup. Subsequent to the verbal introduction, the control group performed the activity lesson without additional instruction. However, before performing the same activity lesson, the treatment group first viewed the 15-minute video presentation of peer models demonstrating the same desired behaviors earlier explained by the teacher to all students.

A list of 32 discrete behaviors was developed by the investigators and represented the dependent variable. Four raters observed the students

and identified whether the desired behaviors were observed (positive) or were not observed (negative). Ninety-two positive and negative behaviors were recorded by each rater for each group. The four raters were unable to record all behaviors; however, interrater reliability coefficients ranging from .89 to .97 were determined by having each rater observe and rate the video model.

Findings

A two-by-two contingency Chi-square test was used to evaluate the total positive and negative scores for the treatment and control groups. The cells contained data computed as a percentage of total observed behavior. Treatment-control versus positive-negative behaviors were found to be dependent ($p < .01$). Dependence due to assigned group was then determined by transforming the percentages to standardized scores for comparison. Twenty-two percent more positive and 22 percent fewer negative behaviors were observed in the treatment students in comparison to the control students resulting in a significant difference ($p < .05$), as predicted.

Interpretations

The investigators concluded that the added videotaped presentation viewed by students in the treatment group increased their chances of acquiring those desired behaviors listed in the rating instrument. Furthermore, their efforts showed that such observational learning studies can be effectively conducted under classroom conditions.

The investigators were careful in identifying the limitations and delimitations of their study. Others were encouraged to conduct similar experiments using larger samples, a wider range of science activity lessons, and improved methods of observing student behaviors.

ABSTRACTOR'S ANALYSIS

The DeTure and Koran article briefly but clearly describes a competent investigation which should be of particular interest to researchers and teachers. This study was one of many successful observational learning inquiries reported in Journal of Research in Science Teaching and is representative of an ongoing line of research being conducted in science education at the University of Florida. Such efforts continue to support the instructional potential of the modeling behavior technique on the acquisition of intellectual process skills by students and teachers. Indeed, it is noteworthy of mention here that Anderson (1973), in his earlier review of science education research, identified another comparative study (Koran, 1971) from the University of Florida series as "illustrative of much of what should be" done, and indicative "of the direction that needs to be taken if comparative studies are to be productive."

The present article contained less than three pages of text and was clearly meant to be a simple description of an exploratory classroom investigation within the larger observational learning domain in science education. Koran (1974) has described this area in detail in a special issue of the Journal of Research in Science Teaching and in a position paper on teacher training (Koran, 1977).

The investigators used a powerful and relevant research design that was rationally linked to the study problem and tested hypothesis. Unfortunately there were no theoretical or experimental works cited in the article, presumably because other studies and reviews in science education, such as those mentioned above, have exhaustively discussed such work. Yet a description of the logical continuity between previous studies and the present study would have been instructive to the general reader. Introductory explanations could have been particularly helpful to those researchers less familiar with this literature.

The description of the study method was reasonable and permits replication of the experiment. Perhaps a small sampling of the observational

items especially abridged for this study would have clarified the nature of the list instrument and the resulting data. Experienced investigators would otherwise have no problems in comprehending the method.

The posttest-only control group design used in this study allowed readers to analyze and interpret the data in an unambiguous, comprehensive manner, thereby encouraging them to investigate other observational learning questions of particular importance to science teachers. The experimental results were powerful and generalizable to a greater extent than suggested by the authors.

Although the apparent goal of the authors was to present a bare-bones explanation of a simple experiment, the brevity of the description did in a few instances leave some questions unanswered in the reviewers' minds:

1. After assigning students to treatment and control group, what was the purpose of further dividing these groups into subgroups? One could assume that the subgroups were simply working units for the particular science experiment being conducted, but the authors state only that behaviors were "recorded for each group" without clarifying whether they meant the large groups or subgroups.

2. High interrater agreement was established from having each rater view the video model and rate it. However, analysis was performed on rating scores for the groups, which would seem more difficult to view and rate. Is such a method of establishing interrater agreement necessarily valid?

3. Finally, we have only the experimenter's opinion that the teacher was "explicit and thorough in her instructions." If the preliminary explanations given by the teacher were inadequate, then the treatment and control groups would have been instruction and non-instruction groups, which would probably account for the significant difference found between the

groups. This does not seem to be the case, but we would have welcomed some discussion and clarification on this point.

These points, however, are not meant to overshadow the otherwise clear and direct manner in which this study was reported. We would have preferred that more clarification had been given to the reader, including an attempt to place this study in perspective with related works in science education or in educational psychology. However, in all other ways, the study was exemplary in its design, execution, and reporting, and represents a valuable addition to our understanding of the role of observational learning as a teaching technique.

REFERENCES

Anderson, R. D. Review of Research in Science Education for the Year 1971. (Research Review Series—Science, Paper 10). Columbus, OH: The Ohio State University, ERIC Information Analysis Center, April, 1973.

Koran, John J., Jr. "A Study of the Effects of Written and Film-Mediated Models on the Acquisition of a Science Teaching Skill by Pre-Service Elementary Teachers." Journal of Research in Science Teaching, 8:45-50, 1971.

Koran, John J., Jr. "Training Science Teachers: Methodological Problems and Issues in Changing Behavior." Journal of Research in Science Teaching, 11:205-218, 1974.

Koran, John J., Jr. "Application of Science Education Research to the Classroom: Validating Science Teacher Training Through Student Performance." Journal of Research in Science Teaching, 14:89-96, 1977.

Hunt, John D. and Bill Lamkin. "An Analysis of Reinforcement and Feedback within an Auto-Tutorial Plant Taxonomy Module." Journal of Research in Science Teaching, 12(2):121-125, 1975.

Descriptors--*Autoinstructional Methods; Biology; Educational Research; *Feedback; *Instruction; Reinforcement; Science Education; Secondary Education; *Secondary School Science; *Teaching Methods

Expanded Abstract and Analysis Prepared Especially for I.S.E. by Gene F. Craven, Oregon State University.

Purpose

The stated purpose of this study was "to compare the performance of students receiving immediate reinforcement and feedback within an auto-tutorial system to the performance of students in the same system but not receiving immediate reinforcement and feedback" (Hunt and Lamkin, 1975). Two null hypotheses were tested using multiple linear regression analysis. First: There is no significant difference between the residual gain scores by the two groups in cognitive knowledge about wild flowers and their parts in reference to the group's variable IQ as measured by the California Test of Mental Maturity. Second: There is no significant difference between the residual gain scores made by two groups in cognitive knowledge about wild flowers and their parts in reference to immediate reinforcement and feedback.

Rationale

The researchers cite professional literature to support their position that there is disagreement relative to the desirability of using immediate reinforcement to maximize cognitive learning. An auto-instructional (audio-tutorial) system was identified as the preferred method of providing the subjects with structured lessons, including immediate feedback and reinforcement to those in the experimental group. Ausubel (1968) was cited as advocating structured lessons because they lead to highly meaningful learning of concepts. Thus, the learning theory of Ausubel was the theoretical basis of the study to the extent

that it supported structured lessons such as those presented via the audio-tutorial (auto-instructional) mode.

No research is cited to support this study nor is a need for the study indicated. Novak (1970) was quoted to support the researchers' position that research related to audio-tutorial methods has been very minimal. They concluded that "gaps that occur in research indicate the need for studying variables within the audio-tutorial paradigm."

Research Design and Procedure

A Pretest-Posttest Control Group Design was used that takes the form:

$$\begin{array}{llll} R & O_1 & X & O_2 \\ R & O_3 & & O_4 \end{array}$$

Two teachers were selected at random from 14 teachers of biology in grades nine and ten of the Waco Independent School District, Waco, Texas. Subjects were picked randomly by the researchers from the class rolls containing the names of 150 students of the two teachers. The first 24 students picked randomly formed the control group, the second 24 formed the experimental group.

The control group consisted of 17 boys and 7 girls; the experimental group included 8 boys and 16 girls. Black, Mexican-American, and white students participated in the study but no information is provided relative to the number of each in either the control or the experimental group.

An auto-tutorial system consisting of 35 mm. slides, slide sequence sheets, glossaries of terms, worksheets, and living plant specimens was complemented by different scripts for the two groups recorded on audio-cassettes. The system was used on a pilot basis with 30 tenth grade students during the 1969-70 school year and revised for the present study. The revised scripts were written so that the student would need no help from the classroom teacher in following the

directions. Scripts used by students in the experimental group included answers to questions and words of praise and encouragement such as "That's right" and "You're doing a great job." Scripts used by the control group included no answers or words of praise and encouragement.

Eight identical systems were produced for each group. All systems were used at one time during the students' scheduled biology period. Tests were given and the auto-tutorial systems were used by the students over a three-week period.

The content pretest and posttest were identical, consisting of 25 multiple-choice item tests which were administered to each student prior to the start of the module and immediately upon its completion. IQs were obtained from the California Test of Mental Maturity which was given during November 1972.

A multiple linear regression analysis was used to test the hypotheses. A full model was employed to treat "the group and IQ identification of each group separately" (Hunt and Lamkin, 1975) and compared to a restricted regression model in which the groups remained separated but the IQ variable was considered in common. A second regression analysis was undertaken where two restricted models were employed and the data compared to determine whether the students' IQ made any difference in scores on the posttest. In one restricted regression model, the groups were separated but the IQ variable of both treatment groups was consolidated. In the comparison regression model, the treatment groups were separated and the IQs of both groups were omitted.

Also, a multiple linear regression analysis was used to determine if there was a relationship between the residual gains in score made by the two groups as a result of the experimental program. A regression model with IQ and group membership separated was used and compared to a model where the IQ and group membership were cojoined.

Findings

In both of the two auto-tutorial systems, one of which employed immediate reinforcement and feedback, the IQ of the students produced a significant (0.05 and 0.01 levels of confidence) difference in cognitive gains. Significantly (0.01 level) greater gains in cognitive learning were produced by the system employing immediate reinforcement and feedback than by the (same) system without immediate reinforcement and feedback.

Interpretations

Based on the data from 48 students, the researchers concluded that the self-instructional system of immediate reinforcement and feedback was more effective for cognitive learning than a system containing no immediate reinforcement and feedback.

ABSTRACTOR'S ANALYSIS

A review of the literature reveals that this study relates to a relatively large number of earlier studies (Anderson, 1971 and 1972; Flynn, 1970; Geis, 1971; Gregory, 1970; Monitor, 1970; Sturges, 1972; Zahorik, 1970) which pertain to the effect of feedback and reinforcement on the cognitive learning of students using programmed instruction and other forms of self-instruction. Unfortunately, Hunt and Lampkin (1975) do not refer to studies such as those cited above to support the need for their study or to show how it relates to a rather extensive body of research on feedback and reinforcement. They cite Novak (1970) who reported that research related to audio-tutorial methods has been very minimal, a fact that is confirmed by a search of the literature. Yet, the primary focus of the present study is on the effect of reinforcement (including praise) and feedback on the cognitive learning of students. The audio-tutorial module, as used in this study, served essentially the same function as does programmed instruction and other auto-instructional systems in that it provided a predetermined

sequencing of learning situations and a controlled schedule of reinforcement and feedback to the learners.

No theoretical basis is presented for the study. Ausubel's learning psychology is cited to support the audio-tutorial system employed in this study because it results in "a structured lesson which Ausubel advocates because it leads to highly meaningful learning of concepts" (Hunt and Lamkin, 1975). But there is no attempt by Hunt and Lamkin to further relate the findings of their study to Ausubel's psychology or to any other theory of learning psychology.

Ausubel (1968) and Novak (1970), both of whom are cited, emphasize the need for research to determine how varying the presentation of advance organizers and sequences of appropriate instruction affects the rate of assimilation of concepts by students, a problem that was not investigated by this study.

The researchers are commended for their use of an experimental research design (Campbell and Stanley, 1963) that was facilitated by the audio-tutorial system which essentially removed the teacher variable in providing learning situations, reinforcement and feedback.

Random selection was employed to form the experimental and control groups and the researchers were careful not to extrapolate their findings beyond the single school district from which the subjects were obtained. While selected randomly, the sex ratios of the two groups differed significantly. The control group consisted of over two-thirds boys (17 of 24) but the situation was reversed for the experimental group that consisted of two-thirds girls (16 of 24). Could sex-related differences such as interest in doing the learning tasks, need for and receptivity to external motivation (written words of praise), and/or tendency to want to please the teacher have influenced the relative performance of the two groups of subjects? Lawlor (1970) found that for second graders the use of verbal rewards which were not congruent with behavior resulted in less efficient problem solving than either a neutral or no-reward situation and that the giving of rewards that were congruent improved the problem solving efficiency of

girls but not boys. Is this true for ninth and tenth grade boys and girls? If so, does it apply to praise and encouragement such as "That's right" and "You're doing a great job" that was included on the audio-tapes for the experimental group (two-thirds girls) used in this study? If so, to what extent did it contribute to the superior performance of the experimental group?

The researchers report that "Black, Mexican-American and white" students participated but do not give the relative number of each in the two groups. If, as is reported by learning psychologists, "the social class membership of a pupil has important implications for his school achievement, his aspirations for academic success, and his achievement in school" (Ausubel, 1968), data supporting the racial equivalence of the experimental and control groups would have increased this reviewer's confidence in the conclusions of this study.

Gregory (1970) found considerable evidence to suggest that children who are external in their locus of control perform better when feedback is supplied by the experimenter as was the case in the present study. This was not true of children who were internal in their locus of control. They responded best in learning situations where they could observe for themselves the correctness of their own responses.

Geis and Chapman (1971) concluded, subsequent to an extensive review and analysis of pertinent research-based literature, that "it is clear that the printed answer or its analog in other media, is not globally and automatically a reinforcer. The weight of evidence from these global studies was that feedback did not enhance learning as measured by immediate posttest scores or by retention in tests." Flynn (1970) reports findings which "imply that response frames in P.I. and the associated feedback do not strengthen content acquisition as measured by multiple-choice achievement tests."

The findings summarized above, some of which are contradictory to one another and to findings of the present study, suggest the need for more definitive studies to test prevailing assumptions about the effects of feedback on the cognitive learning of students. Less global studies

that better control for variables such as sex, socioeconomic, and psychological factors such as locus of control in addition to feedback schedule appear warranted. The need for further studies on the effect of IQ on factors such as residual gain on posttests is questionable because an extensive body of research shows that IQ is related to practically every cognitive variable that is studied.

REFERENCES

Anderson, R. C. "Conditions Under Which Feedback Facilitates Learning from Programmed Lessons." Journal of Educational Psychology, 63:186-188, 1972.

Anderson, R. C. "Feedback Procedures in Programmed Instruction." Journal of Educational Psychology, 72:148-156, 1971.

Ausubel, D. P. Educational Psychology. New York: Holt, Rinehart, and Winston, Inc., 1968.

Campbell, D. T. and J. C. Stanley. "Experimental and Quasi-Experimental Designs for Research on Teaching" in N. L. Gage (ed.) Handbook of Research on Teaching. Chicago: Rand McNally & Company, 1963.

Flynn, M. J. "Effects of Different Frequencies of Response Frames in Programmed Instruction." American Educational Research Journal, 7:569-581, 1970.

Geis, G. S. and R. Chapman. "Knowledge of Results and Other Reinforcers in Self-Instructional Systems." Educational Technology, 11:38-51, 1971.

Gregory, M. K. "Effects of Locus of Control and Type of Reinforcement on Programmed Instruction Performance of Adolescent Boys." Journal of Educational Research, 70:45-49, 1976.

Lawlor, F. X. "The Effects of Verbal Reward on Behavior of Children in the Primary Grades of a Cognitive Task Typical of the New Elementary Science Curricula." Journal of Research in Science Teaching, 7:327-340, 1970.

Monitor, K. "Affects of Using a Self-Scoring Sheet on Knowledge Retention." Journal of Educational Research, 63:437-439, 1970.

Novak, J. D. "Relevant Research on Audio-Tutorial Methods." School Science and Mathematics, 70:777-784, 1970.

Sturges, P. T. "Effect of Instructions and Form of Informative Feedback on Retention of Meaningful Material." Journal of Educational Psychology, 63:99-102, 1970.

Zahorik, J. A. "Teacher Verbal Feedback and Content Development." Journal of Educational Research, 63:419-423, 1970.

Kren, Sandra R. and John P. Huntsberger. "Should Science Be Used to Teach Mathematical Skills?" Journal of Research in Science Teaching, 14(6):557-561, 1977.

Descriptors--Achievement; Educational Research; Elementary Education; Elementary School Science; Instruction; *Mathematics Education; Science Course Improvement Projects; Science Education; *Skill Development; *Teaching Methods; *Transfer of Training

Expanded Abstract and Analysis Prepared Especially for I.S.E. by George W. Bright, Northern Illinois University.

Purpose

The purpose was to determine whether mathematical skills can be acquired by presenting either (1) quantitative science exercises alone or (2) science and mathematics exercises concurrently.

Rationale

A response to "demands to return to the basics" prompted an application of the "several theories of transfer of training" to science teaching in terms of the effects on mathematics achievement. Integration of the curriculum is presented as one way to achieve transfer. (It is not clear whether the references cited concerning attempts to integrate science and mathematics are reports of research or exhortations to integrate mathematics and science. The common element reported from these articles is the authors' beliefs that mathematics and science should be integrated.) The study was viewed as a follow-up to a study by Kolb (1968) in which greater achievement of science objectives was observed when a mathematics sequence preceded the science exercises.

Research Design and Procedure

The science exercises were from Science: A Process Approach I (prediction in various physical systems, and the measurement of angles). The mathematics instruction was lecture-demonstration based on a textbook supplemented by a packet of worksheets "to insure adequate coverage"

(interpretation and construction of linear graphs; measurement and construction of angles).

The treatments were (1) science, (2) mathematics and science concurrently, (3) mathematics and (4) control (pre- and posttests only, with a 10-day lapse). The treatment lasted "approximately 12 consecutive school days."

Students were 161 fourth- and fifth-graders from eight classrooms. "Three classrooms were randomly assigned to treatment groups and five classrooms volunteered for one of the treatments used in the study" [emphasis added]. Distribution of classrooms among the treatments was not reported. The experimental design was a non-equivalent control group design.

Prior to the treatment, Sections I and II of the Kren Test were given. The test was administered again the fourth day after the treatment. Analysis of covariance, multiple range test, analysis of variance, multiple classification analysis, correlation, and item analysis frequency were used.

The Kren Test-Section I is on measuring and constructing angles; the Kren Test-Section II is on interpreting and constructing linear graphs. Each section consists of 15 items; the total reliability was .93. Content validity was endorsed by a panel of readers.

Findings:

ANCOVA (the covariate is not clearly identified) was used on posttest scores to measure differences among groups. For each section of the test, significant differences ($p < .01$) are reported among the treatment groups. The df for each F-statistic is given as (3,157).

Mathematics-only and mathematics-science-concurrently were "equally effective" in teaching the material on angles, but science-only was "not an acceptable substitute." "Analysis of covariance was used to analyze the data in order to adjust for initial difference [among]

groups, which might have been present due to varying amounts of ... to the prerequisite skills." The material on graphs was ... with equal effectiveness via mathematics-only, science-only, or mathematics-and-science-concurrently.

ANOVA was used on posttest scores to measure grade effect. A significant effect ($p < .01$) was reported only for Section I. Since means are not reported, it is impossible to tell which grade scored higher. Also, the df for the analysis is not reported.

For each section of the Kren Test, there was a positive correlation between pre- and posttest scores. There was also a positive correlation between Section I and Section II for the posttest only. Neither the correlation nor the significance levels were reported.

Constructing angles was harder at both grades than reading a protractor or measuring an angle. No statistics are reported for the multiple range test or the multiple classification analysis.

Interpretations

The science activity on prediction seems improperly placed. It should be delayed. Further investigations of the effectiveness of the integration of science and mathematics should be undertaken.

ABSTRACTOR'S ANALYSIS

The study has obvious flaws. Of primary importance is that since classes were assigned to treatments (but not randomly) the unit of analysis should be the class mean. The correct df for the ANCOVA and ANOVA are (3,4) and (1,6) respectively. It is impossible to tell whether a reanalysis would result in any significant F-statistics.

The comments below are less critical.

1. The rationale for investigating the "integration of the

curriculum" as part of the transfer of training pardigm is not clear. It is, of course, one aspect of the learning environment, but there is no justification that it is a critical aspect.

2. Twelve consecutive school days is long enough to anticipate some results, but the description is "approximately 12 days." Were the treatments of different lengths?

3. The nature of the use of the mathematics concepts in the science instructional treatment should be explained. This seems especially important if we are to learn to recognize the kinds of integration of curricula that are effective.

4. What did the control group do for 10 days? Why was the pre-test/posttest lapse 10 days rather than 12?

5. What was the distribution of grades among the treatments?

6. The results are reported in a confused way. Some are in the discussion section but not the results section. Results for two analyses seem not to be reported at all.

7. The F-statistics alone are impossible to interpret. Means and multiple range test statistics are essential for the interpretation.

8. Differences in exposure to prerequisites among the treatment groups are cited as the cause for using ANCOVA to measure the treatment effect. It seems logical that there would also be differences in exposure to prerequisites between the grades. Why wasn't ANCOVA (instead of ANOVA) used to measure the grade effect?

As a whole, the study as reported lacks credence. The reader cannot determine from the information provided whether science can in fact be used to teach mathematical skills.

Valeriote, I. M. "A Self-Paced Laboratory Course in First-Year General Chemistry." Journal of Chemical Education, 53(2): 106-108, 1976.

Descriptors--Autoinstructional Methods; *Chemistry; *College Science; Educational Research; Higher Education; *Laboratory Procedures; Science Education; Self Help Programs; *Teaching Methods

Expanded Abstract and Analysis Prepared Especially for I.S.E. by J. Dudley Herron, Purdue University.

Purpose

The purpose of the investigation was to test the feasibility of a self-paced laboratory course in a first-year college chemistry course. Although comparisons were made between laboratory sections that were self-paced and those that were not, no specific hypotheses were tested.

Rationale

No previous research is cited as a basis for this study. The rationale for testing the self-paced laboratory is that little research of this nature has been done in spite of the general opinion that students in first-year college courses differ markedly in motivation, learning ability, and background.

Research Design and Procedure

One might reasonably debate whether a true research design existed for this study, and the author appears somewhat ambivalent about the matter. In the conclusions the author states: "Comparisons between groups have been made but the self-paced course was not instituted to try to prove that the students involved perform significantly better than regular students." However, in an earlier section the author speaks of "the post-test only control group design" which suggests that some kind of significance test was planned.

The design is best described as a comparison of intact groups. Two laboratory sections of 24 students each were selected from a total of about 17 sections. (The number of sections is not given but the total number of students is given as approximately 400.) Students in these sections were informed that they would become a "trial or special group in that they would not be required to follow a strict laboratory schedule."

In addition to the regularly-scheduled, three-hour laboratory session afforded to all students, the self-paced students were given an opportunity to come to laboratory during an additional 18 hours each week. (The author does not state how much students used this additional time.) They could work at their own pace to complete five set experiments each term and were given extra credit for optional experiments selected from a "menu." Regular students were required to complete the five set experiments plus three menu experiments during the first term and five during the second term.

It should be noted that self-pacing was not the only variable in the experiment. Students in the self-paced sections were required to write a report for each experiment, and this report was graded in class at the time it was turned in. A new experiment could not be started until this grading was done. By contrast, the regular students were required to write a formal report for only two experiments each term. In addition, students in the self-paced course completed a pre-test over each lab and grading criteria were different for the two groups. In fact, it would appear that the only things that the two groups had in common were the five set experiments and laboratory examinations given at the end of each term.

The author reports data on the number of units completed each month, the mean score on the laboratory examinations, the final laboratory grades, and student responses to an opinionnaire concerning the course.

Findings

Students in the regular course completed slightly more laboratory units than did the self-paced students (a mean of 20 units vs. 18.5 units.)[1]

The self-paced students had a higher mean for the laboratory examination given at the end of each term and they also had a higher mean for the final laboratory grade (see Table 1). The results on the opinionnaire suggest that students a) would like to take other self-paced courses, b) liked the idea of having reports graded in their presence, c) liked the grading system that depended mainly on the number of experiments completed, and d) felt they learned more in this course than in other laboratory courses. Other questions on the opinionnaire dealt primarily with mechanics of the course and were not truly evaluative in nature.

*Table 1. Laboratory Examination Grades and Final Laboratory Grades for Regular and Self-Paced Groups

	Mean Percentage		
	1st Term Laboratory Examination	2nd Term Laboratory Examination	Final Laboratory Grade
Regular group	49	68	65
Self-paced group	52	79	72

*Table 3 in original.

[1]It isn't entirely clear what constitutes a "unit." The reviewer presumes that some laboratory exercises were more involved and counted as more than one unit.

Interpretations

The author clearly believes that the self-paced course is at least as good as the regular course and that it is better for some students. He states that "it is hoped that the self-paced laboratory course will be integrated with a modular self-paced theory course." He appears to see this as an option to a regularly scheduled course rather than a replacement.

ABSTRACTOR'S ANALYSIS

When reading this report, one has the feeling that the author had decided that a self-paced laboratory would be a good thing before doing the investigation. It is not clear why the facade of an experimental design was incorporated in the report. There are numerous alternative hypotheses to explain the results and there are occasional hints that the author was aware of this. Perhaps the author needed to convince an administrator that the self-paced lab would be a good thing and proceeded to carry out the necessary arm-waving.

For the benefit of those readers who may be looking for the definitive study in this area, we mention a few of the reasons that the results could be spurious.

The author did not randomly assign students to the treatments and he provides little evidence that the class sections selected for treatment are like other sections in all important respects. The author comments that "although the pretest is a concept upon which research workers rely to ensure equality of experimental and control groups before the differential treatments, it is not actually necessary for an experimental design. The most adequate way of preventing initial biases is randomization..." This comment is well taken and we can think of no pretest that could have been used in this study to convince the reader that the groups were comparable in all of the variables that might have affected the outcome of this investigation.

However, we are still left with the fact that we have no real reason to believe (or not believe, if you wish) that the two comparison groups constituted random samples drawn from the same population.

Presumably the author was interested in the effect of self-pacing, but the treatment involved far more than that. The author suggests that the self-paced students scored higher on the laboratory examinations because "(1) they had to write up each experiment, (2) the trading of each experiment was a learning experience on a one-to-one basis, and (3) they had the time available to do the experiments more carefully." None of these conditions require self-pacing.

Students in the self-paced group were told that they were a special group. Based on my review of educational research, there is no surer way to increase the performance of a group of students than to convince them that they are receiving special attention. The so-called Hawthorne effect seems to be very powerful and could easily account for the results in this study.

Although I believe that any of the above factors could account for the results obtained, I am not saying that they do. It is just that the study does not enable one to attribute the observed effect to a specific cause.

The author's references to "random selection of subjects" and "post-test only control group design" are misleading. They imply that a controlled study was conducted when it was not. It would have been better had the author limited the paper to a frank description of the two treatments, a report of any objective data that were available, and a discussion of the author's subjective judgments concerning the value of the innovation. Over the years, such reports (without benefit of experimental design) have undoubtedly influenced educational practice more than has educational research.

Most practitioners are skeptical of educational research and for good reason. A great deal of what is represented as research is similar

to this study where numerous competing hypotheses could explain the data that are reported. "Research" implies a kind of rigor that is absent and begs for confidence in results which is often unjustified. When later observations under different conditions contradict the original "research," the practitioner is left confused and skeptical.

If we ever hope to establish credibility, we must be totally honest with ourselves and with our audience. When we believe that we have strong evidence, we should report it as such. When we have hunches and opinions, we should feel free to share them, but we should be clear about what they represent.

With this preamble, I offer some hunches and opinions concerning this study and the general issue of self-pacing in the laboratory.

I see no reason to doubt the reported differences in performance on the laboratory examinations. I also believe that the author's explanation for these differences is quite plausible. He suggests that "the self-paced students scored higher on the examinations because (1) they had to write up each experiment, (2) the grading of each experiment was a learning experience on a one-to-one basis, and (3) they had the time available to do the experiments more carefully." I would add to this the suggestion that the pre-lab assignment, followed by a pre-lab test which had to be passed before commencing the experiment, also contributed to the difference in performance. These opinions are not based on hard data reported in this study. However, they are consistent with general principles established by years of educational research and practice.

People do better at activities which they prepare for. By giving the self-paced students a pre-lab assignment, they made advance preparation for their work. By requiring them to pass a pre-lab examination before they began the lab, there was some assurance that they actually did the advance preparation. If they did not, there was sure to be an opportunity for clarification by the instructor or review by the student.

The more mentally involved students are in an activity, the more they learn. Certainly the requirement of writing a report of the laboratory investigation required the student to think about the laboratory activity. Thus, the self-paced students were required to devote some attention to each laboratory activity on three separate occasions: (1) during the pre-lab assignment, (2) during the actual experiment, and (3) while writing the report. One might add as a fourth occasion, the time that the report was being graded in their presence. Based on the article, it appears that students in regular classes were required to deal with the laboratory activity only during the actual experiment.

Knowledge of results or "feedback" has been established as important for learning. The self-paced students received useful feedback on two occasions: first, they learned whether they understood what was to be done in the lab when they took the pre-lab exam, and second, they learned how well they understood the actual experiment when their report was graded. It should be noted that this feedback was provided immediately as the report was graded and that students in the regular sections did not receive such feedback.

Although research is not extensive in this area, there are studies which indicate that careful plodding to the point of understanding a smaller amount of material results in more learning than covering a larger block of content at a pace that exceeds the rate of comprehension.[2] This is the whole notion behind Bloom's Mastery Learning strategy. Therefore, it is plausible that students in the regular laboratory program may have done poorly because they were forced to rush through some investigations in order to finish in the allotted time. The self-paced students did not labor under this handicap.

None of the conditions that appear to be important to the success of this laboratory program require self-pacing, even though an arrangement that allows for self-pacing may make it easier to satisfy those

---

[2]See for example, Gabel, D. and J. D. Herron. "The Effects of Grouping and Pacing on Learning Rates, Attitude, and Retention in ISCS Classrooms." Journal of Research in Science Teaching, 14(5): 385-399, 1977.

conditions. From my reading of research, the format under which a course is offered is generally less important than factors such as (1) clearly communicating to students what they are expected to learn, (2) providing students with a rich learning environment which is compatible with their level of intellectual development and their prior experience, (3) providing students with clear and frequent feedback concerning progress toward course objectives, (4) not forcing the student to move on to a new topic until they have mastered current topics, and (5) maintaining a learning environment that is sufficiently interesting that the student persists at the learning task. These appear to be important variables operating in this study as well.

Wideen, Marvin F. and Richard L. Butt. "Student and Instructor Directed Approaches to Elementary Science Education: A Comparative Analysis." Journal of Research in Science Teaching, 12(2):95-100, 1975.

Descriptors--College Science; Educational Research; *Elementary School Teachers; Higher Education; Methods Courses; *Preservice Education; *Science Education; *Teacher Education; *Teaching Methods

Expanded Abstract and Analysis Prepared Especially for I.S.E. by John D. Balling, Chesapeake Bay Center for Environmental Studies.

Purpose

The major purpose of this study was to compare student directed and instructor directed approaches to teaching at the college level, within the context of an elementary science education course. A secondary goal was to determine how variation of the course content interacted with the two teaching methods.

Rationale

As a rationale for the study, the authors appeal to what they say is a very general concern in education today, namely the efficacy of self-determination in professional education courses. No previous research is cited as suggesting this study and no specific models or theories are mentioned.

Research Design and Procedure

Most of the 76 undergraduate students (69 female and 7 male) comprising the initial sample for this study were in their second year of teacher preparation at the University of Saskatchewan, Regina Campus. Each subject was enrolled in one of two sections of an introductory science methods class. One section, labeled the primary option, was designed for students interested in grades K-3, and the other, called the intermediate option, was for students more interested in teaching grades 4-8. The G.P.A. for the students in the primary option section was reported as higher than that

for the students in the intermediate option section. No indication is given as to either what statistic was used to determine the relative G.P.A.'s or how much higher the G.P.A. of the primary section actually was. Due to attrition for various common reasons, the authors state that the final data analysis included 66 subjects. However, the numbers of subjects per group reported in a footnote add up to 61. No other information about the subjects is provided.

The goals of the course for the two sections were identical and included both process and content objectives. A wide variety of materials were made available to the students, and they had contact with children each week in a micro-teaching laboratory situation. Grading for both sections was done on a contractual system. A major difference between the two sections was that students taking the primary option were involved in a three-week student teaching session, while the students pursuing the intermediate option had no analogous experience.

The actual experimental treatments were described very sketchily although the authors stated that they would provide more information on request. "The student directed approach (SDA) involved a high degree of decision making for the student and a supportive, consultative role for the instructor. . . . The instructor directed approach (IDA) consisted of the more conventional pattern of lecturers and workshops" (Wideen and Butt, p. 96). The SDA group used self-instructional materials and pursued these activities on their own time, while the IDA group met at regularly scheduled times. A single instructor, a student assistant, and a laboratory aide handled all the teaching for both the SDA and IDA groups in both the primary and intermediate option sections.

Four dependent measures—a process measure, a cognitive learning test, an attitude scale, and a rating of an end-of-term interview—were used to assess the treatment effects. The process measure consisted of items selected from Form A, Science Process Measure published by AAAS. No reliability estimate for these items was provided, although the authors do state that an inter-scorer consistency of 0.88 had been previously obtained. The 43-item cognitive test had a reliability of .89, and the reliability of the 8-item attitude scale was .78. In an interview at the

end of the course, the instructor gave each student an overall rating on a five-point scale. As a manipulation check, a second instructor observed the treatment groups to determine whether the behavior of the students in each group actually differed. How these observations were made and whether they were in any way systematic is not mentioned.

The two major independent variables were treatment group (SDA or IDA) and orientation or class section (primary or intermediate). The students in both the primary option and the intermediate option sections of the science methods class were randomly assigned to one of the treatment groups. Thus, a basic entirely between-subject 2 x 2 factorial design with unequal cell frequencies resulted, and the data on each of the four dependent measures were analyzed within this framework.

Findings

1. The authors state that the observational procedures revealed that the behavior of the students in the two treatment groups was different. However, the observer was not blind to the manipulation and no statistical support for the importance or magnitude of this difference is provided.

2. No significant main effects for the treatment (SDA versus IDA) were found on any of the dependent measures (process, cognitive, attitude, or interview).

3. Students in the intermediate section scored significantly higher than did those in the primary section on the cognitive measure.

4. Statistically significant Treatment by Section interactions were found for the attitude and interview measures. In both of these interactions, the primary group scored higher under the SDA than they did under the IDA, while the reverse was true for the intermediate group. A similar interaction was obtained on the cognitive measure but it was significant only at the 0.10 level.

Interpretations

1. As Wideen and Butt note, any conclusions drawn from this study are tentative because of: 1) the limited population from which the sample was drawn; 2) the absence of pretests to determine the equality of the treatment groups on the relevant measures; 3) the fact that subject variables, such as sex and academic background, that could have interacted with the treatment, were not included in the design, and 4) the presence of an interaction between treatment and class section for the attitude and interview data which presents problems for the generalization of the results.

2. The learning of process skills was similar for both the SDA and IDA groups. However, within the SDA group, a great deal of individual variation was observed in the use of the process modules. Some students studied them diligently throughout the term, others hardly at all. The authors suggest that since no difference was found between the two approaches, that a substantial amount of freedom could be given to science education students for learning process skills.

3. Wideen and Butt suggest that the interaction obtained on the attitude and interview measures in which the primary option section subjects did better with the SDA than with the IDA while the reverse was true for the intermediate option students could be due to initial group bias, the selective transfer of task relevant skills from other classes and from student teaching, or opposite patterns of preference for the two modes of teaching in the two sections. None of these explanations could be tested within the constraints of the study.

4. The results tentatively suggest that neither the SDA or the IDA has any overall superiority in the attainment of course objectives. However, given the interaction of the treatment with the class option selected by the students, it may be that different students benefit differentially from different types of instruction.

ABSTRACTOR'S ANALYSIS

Particularly in relatively new areas of research in which the methods and dependent measures are not well standardized (most science education research falls in this category), the general absence of treatment effects provides virtually no information. In the Wideen and Butt study, many interpretations of the null result are possible. It could be that, as far as the students are concerned, there really are no important differences between instructor-directed and student-directed approaches to teaching science education. However, the study, as designed, can present no new insights as to why such a state of affairs might be true. No theory which predicts the absence of difference between the IDA and the SDA was discussed. It could also be that there are real differences between IDA and SDA, but not on the dependent measures used in this study. For example, Semb, Conyers, Spencer, and Sosa (1974) have found that a manipulation similar to the SDA used by Wideen and Butt tends to increase course withdrawal rates as compared to an IDA condition.

Another possibility is that there may be actual differences on the dimensions sampled by the authors (acquisition of science process knowledge, cognitive learning, attitude change), but the measures used were simply too insensitive to detect them. Finally, some unmeasured subject variable, e.g. academic background, could be over-riding any treatment effects. All of these hypotheses, plus innumerable others, could easily account for the data.

In the absence of any theoretical notions, the treatment by section interactions found on the attitude and interview measures are also difficult to interpret. Apparently, the two different class sections were included to increase sample size and for reasons of convenience to the experimenters. The authors probably hoped that there would be no effects due to class section and certainly no interactions with the treatment. The findings regarding class section in this study underscore a general problem in educational research. In order to insure an adequate sample size, it is often necessary to include different classes, teachers, schools, or school districts in the design of the experiment with the hopes that these "control" factors will contribute relatively little to the overall variance in the dependent measures and that they will not interact with

the independent variables explicitly manipulated in the study. When the "control" factors do appear to be important, it is quite often difficult to interpret the results and a general appeal is made to some unspecified individual difference phenomenon, whether it be differences among students, teachers, or schools. Although there is no easy solution to this problem, enough levels of these "control" factors should be sampled so as to insure their adequate representation and a reasonable sampling distribution. The adequacy of representation depends on the population about which statistical generalizations are to be made. Decisions about the population to be sampled are unique to individual projects and depend upon the goals of the experimenter. The necessary properties of the sampling distribution depend upon the variance in the dependent measure and the magnitude of the treatment effects the experimenter wishes to detect. By reviewing previous research in the area, an experimenter can often get some feel for how large a treatment effect can be expected and for how much variance is contributed by using different classrooms, schools, etc. The number of levels of these "control" factors necessary to insure adequate power to detect the mean differences produced by the treatment could then be calculated (see Cohen, 1969).

It might also be helpful to consider an experiment as a series of independent replications across one of these "control" factors, such as classes, the statement that SDA was more effective than IDA in one class but that the reverse was true in the other is not very helpful. However, if the treatment had been replicated in more classes, and a statement such as in four out of five classes the SDA produced more cognitive learning than the IDA could be made, then a great deal of information about the treatment is available. Further, in the situation with a large number of classes, it might be easier to determine what was different about the aberrant class(es).

As the authors point out in their discussion, the validity and generalizability of this study are in question for various reasons. No theoretical analysis of the role of student self-determination in education was provided either to guide or interpret the study. As a result, the conceptual contribution of this study to the understanding of the difference between student-directed and instructor-directed approaches to teaching is quite

minimal. The study also makes no new methodological contributions. A conventional design and relatively common dependent measures were used. Although univariate analysis of variance procedures were used to analyze the data, the entirely between-groups factorial design of this study presented an excellent opportunity for the use of multivariate analysis of variance. Use of MANOVA would have provided protection against the effects of correlations among the dependent measures (they were not reported) and against the tendency for certain tests to be significant by chance as more variables are included in the analysis.

The experimental design employed by Wideen and Butt was rather weak. There were a great many differences between the two treatment groups; that is, the SDA and the IDA were different on many dimensions. As a result, any causal inferences were impossible. Even if treatment effects had been found, there would have been no way to know which differences between the SDA and the IDA were important, and which were not. A better experimental design would systematically vary particular aspects of the teaching situation, holding others constant. For example, all groups could be given the same learning modules but one group would be allowed complete self-pacing, another group would be rigidly instructor paced, and a third group would be allowed self-pacing within instructor defined limits. In this way, the effect of student control over pacing per se could be analyzed.

The written report provided by Wideen and Butt is rather vague. The study needs a stronger introduction giving more background and the rationale for this research project. The methodology section was also greatly abbreviated. The details of the two treatments, SDA and IDA, are not reported, although the authors say they will provide them on request. For the sake of clear scientific communication, the partial report of experimental methodology should be discouraged, but the concise report of methods should not.

Finally, the authors do not specifically relate their study to other research concerning student self-determination in higher education. It would be nice to know exactly how Wideen and Butt feel their study fits into the existing literature on personalized systems of instruction (PSI).

Although relatively little research has been done on the use of PSI in elementary science teacher education, there is no reason to expect greatly different results when the course content is changed from one academic subject (for example, psychology which has been heavily researched) to another (science methods). Similar studies with more controlled variables have been done previously (see Johnston, 1974; Sherman, 1974). Future research in science education concerned with student self-determination could borrow much more heavily from the existing PSI literature.

REFERENCES

Cohen, J. Statistical Power Analysis for the Behavioral Sciences. New York: Academic Press, 1969.

Johnston, J. M. (Ed.). Behavior Research and Technology in Higher Education. Springfield, ILL: Charles C. Thomas, 1974.

Semb, G.; D. Conyers; R. Spencer; and J.J.S. Sosa. "An Experimental Comparison of Four Pacing Contingencies in a Personalized Instruction Course." In J.M. Johnston (Ed.) Behavior Research and Technology in Higher Education. Springfield, ILL: Charles C. Thomas, 1974.

Sherman, J. G. (Ed.). Personalized System of Instruction: 41 Germinal Papers. Menlo Park, CA: Benjamin, 1974.

PIAGETIAN STUDIES

Lawson, A. E. and J. W. Renner. "A Quantitative Analysis of Responses to Piagetian Tasks and Its Implications for Curriculum." Science Education, 58(4):454-559, 1974.

Descriptors--Curriculum; College Students; *Educational Research; *Intellectual Development; *Learning Theories; Measurement; Science Education; Secondary School Science; *Task Analysis

Expanded Abstract and Analysis Prepared Especially for I.S.E. by Fred W. Finley, University of Wisconsin-Madison.

Purpose

Lawson and Renner stated the following questions:

1. Can a battery of Piagetian-styled tasks be used and scored quantitatively to determine what percentage of the school population in grades 7 through 12 (ages 12 through 20 years) has entered the stage of formal thought?

2. Can Piagetian tasks be analyzed quantitatively to determine whether or not they measure the psychological parameters of concrete or formal operational thought they purport to measure? (p. 549).

The implicit hypothesis of the second question is that if a set of students is tested with a battery of Piagetian tasks, then the relationship among the scores will indicate two factors: one representing concrete and the other representing formal operational thought. Each task thought to measure concrete operations and each task thought to measure formal operations was expected to be related to one factor but not the other. The first question actually was dependent upon finding an affirmative answer to the second question.

Rationale

Piaget's model of intellectual development provided the framework for the study. Piaget has proposed four stages of intellectual development,

the last two of which (concrete and formal operational) were central to the study. The principal characteristics of the concrete operational stage are that the child is tied to particular situations and objects and that mental operations in one area may not be applied to another. The formal operational child, on the other hand, can operate on symbols, recognize the logical inadequacies in an argument, and apply the same logical operations to different tasks which require the same logic. The general age guidelines for concrete and formal operational thought are 7 to 11-12 years and 11 to 15 years respectively (Piaget, 1966). However, Lawson and Renner indicated that a number of studies in this country show formal thinking develops later in children or may not develop at all.

Research Design and Procedure

Three separate studies were reported. They were different in terms of the population of subjects and the set of Piagetian tasks given the students. The research design and statistical procedures for each study were identical. Students were randomly selected and interviewed as they performed a set of Piagetian-type tasks. Scores were assigned during each task to indicate the developmental level of students. A total score was labeled as representing concrete operational, post-concrete operational, and formal operational stages. The first question was answered using percentages of the samples which were found to be at each of the three levels. The second question was answered on the basis of a principal component analysis.

The sample for the first study included 588 students randomly selected from 25 schools. The median age was 14.8 years. Six Piagetian tasks were used. The first three involved the Conservation of Solid Amount, the Conservation of Weight and the Conservation of Volume when the shape of a clay object is changed. The other tasks involved the Conservation of Volume by water displacement when the masses of objects were different, the Elimination of a Contradiction when a large heavy block floated while a small light block sank, and the exclusion of Relevant Variables using a pendulum.

The sample for the second study was 143 students randomly selected from 300 college freshmen attending a private Oklahoma university. The median age was 18.6 years. The Conservation of Volume task using clay, the Elimination of Contradictions, and the Exclusion of Relevant Variables tasks were used as in the first study. In addition, a Reciprocal Implications task which involved the discovery of the relation between the angle of incidence and angle of reflection for an object striking a surface and a Separation of Variables task using six flexible steel rods were included.

The samples for the third study were 51 biology students and 50 chemistry students who were randomly selected from a single large high school. All 33 physics students were included. The mean ages were 16.4 years, 17.3 years, and 17.9 years, respectively.

Four tasks were administered to each student, three from the previous studies: Conservation of Weight, Conservation of Volume using cylinders, and Separation of Variables. The fourth task was an Equilibrium in the Balance task using a balanced beam and hanging weight. In addition, two written tasks, a Ratio Problem and an Islands Puzzle, were administered to groups of 15 to 50 students.

Findings

The findings of each of the three studies with respect to the distribution of students between concrete and formal operational stages were as follows:

Study 1: The percentages of students found to be in concrete operations for grades 7 through 12 were 83 percent, 77 percent, 82 percent, 73 percent, 71 percent and 66 percent, respectively. Few students in each group were in the formal operational stage (1 to 12 percent).

Study 2: The percentages of college freshmen whose scores reflected concrete operational, post-concrete operational, and

formal operational thought were 51 percent, 27 percent, and 22 percent.

Study 3: The percentage of concrete operational studies was 29.5 percent in biology, 40 percent in chemistry, and 3.0 percent in physics. The percentages of students in post-concrete operations were 35.3 percent in biology, 18.0 percent in chemistry, and 33.3 percent in physics. The percentages of students in formal operations were 35.2 percent in biology, 78.0 percent in chemistry, and 63.7 percent in physics.

The second question was whether or not the scores for the set of Piagetian styled tasks reflect the theory upon which they were based. The results from each study with respect to this question are summarized below.

Study 1 - Seventh Through Twelfth Grade Students. The six tasks in the study were predicted to measure two psychological factors. Two corresponding components accounting for 31.25 percent and 23.84 percent of the variance were determined to be significant. The Conservation of Solid Amount and Conservation of Weight task which purported to test for concrete operations correlated with the second component at 0.7758 and 0.7895. The Conservation of Volume using cylinders, Elimination of Contradictions and Exclusions tasks correlated with the first factor at 0.7522, 0.7429 and 0.7299, respectively. The fourth task, Conservation of Volume, correlated moderately with the first component (0.4544) and the second component (0.4402).

Study 2 - College Freshmen. All tasks used on this study except for the Conservation of Volume using clay were designed to test formal operational thought patterns. One component was identified which accounted for 48.3 percent of the variance. The correlation between the Conservation of Volume task and the first component was low (0.2547). The other tasks all loaded heavily on the second component. The component loadings for each task were: Reciprocal Implication, 0.7081; Elimination of Contradictions, 0.8053; Separation of Variables, 0.8068; Exclusion, 0.6818.

Study 3 - High School Biology, Chemistry, and Physics Students. The principal component analysis extracted one component which accounted for 62.2 percent of the variance. The Conservation of Volume, Separation of Variables, Equilibrium in the Balance and Ratio tasks all correlated highly with this component (0.84, 0.85 and 0.80, respectively). The other two tasks, Conservation of Weight and the Islands Puzzle also loaded substantially (0.74 and 0.69) but the common factor variances (55.4 percent and 47.97 percent) were less than for the other tasks which range from 63.36 percent to 72.13 percent.

Interpretations

With respect of the substantive question concerning the ages at which students can be expected to be formal operational, the authors state that "far fewer subjects exhibited the formal operational thinking than expected on the basis of the 12 to 15 years of age figure" (p. 557). They suggest several possible explanations. One of these was that students have not had educational experiences which would promote formal reasoning. Another was that the tasks involve specific content from physics. The authors clearly favor the first explanation. With respect to the methodological question the authors conclude that the sets of tasks used in the three studies do reflect the two components predicted by Piaget's theory—concrete and formal operations. Although not explicitly stated, this implies that the authors were convinced that the quantitative scoring of the Piagetian-styled tasks was effective in assessing the developmental stage of groups of students.

Data supported the claim that the Conservation of Solid Amount task measured concrete operational characteristics and that the Conservation of Volume using cylinders, the Elimination of Contradictions, the Exclusion of Irrelevant Variables, Reciprocal Implications, and Separation of Variables, Equilibrium in Balance, and Ratio Problem tasks measured formal operational characteristics. The factor loadings for the Conservation of Volume using clay task (Study 1) and Piaget's description of the task implied that both formal and concrete operational characteristics were involved. The factor loadings for

the Conservation of Weight task (Studies 1 and 3) and for the Islands Puzzle task (Study 3) suggested formal operational characteristics were being measured, but the results were not entirely convincing to the authors.

ABSTRACTOR'S ANALYSIS

There are two important contributions of the study. First, the study provides evidence that the age norms for concrete and formal operational stages reported by Piaget are not consistent with those found in the Lawson and Renner sample. The study shows the proportions of students at formal operations to be far less than would be expected. The second contribution is methodological. The study provides evidence that individually administered Piagetian tasks can be scored quantitatively and that the resulting data can be analyzed meaningfully. Descriptions from previous studies had been more qualitative and accordingly somewhat more difficult to interpret in testing large numbers of students. There was extensive agreement between what was expected on the basis of the theory and what was found. The theory predicted concrete and formal factors for certain tasks and results were generally consistent. Additional comment on the implications of this methodological gain would have been useful.

The research design of the studies was adequate. The design allowed the researcher to answer the questions of interest. The statistical procedures were appropriate for the questions and design. In particular, the use of principal component analysis techniques proved to be a powerful technique for assessing the underlying structure of empirical evidence. The only information on research procedures not reported was the criteria used to assign the numerical scores for the performance of each task. Explicit statements of these criteria would be necessary to anyone who would replicate the study or use the procedures for other purposes.

The interpretation of the findings is consistent with the Piagetian concern about the logical form of students' thought. However, it

should be noted that this does not address, and in fact is not intended to address, questions regarding how students use their knowledge of science laws and concepts. This is evident in the authors' argument that the tasks are content-free. Presumably this statement is intended to mean it is not necessary to know physics to perform the tasks accurately. However, it should not be understood to mean students do not use their knowledge of physics. Stating a task is content-free is different from stating that student thinking is content-free. What students are required to do in a task may be very different from their thinking. It is likely that the character and quality of performance will be determined by both the development of logical forms of thought and the learning of concepts and laws relevant to the tasks.

To the extent student knowledge of specific content interacts with the development of thought, one of the educational implications of the authors can be questioned. The proposal that if a majority of the secondary school students is concrete operational and the content is formal operational then the curriculum "should be reshaped to reflect concrete objects, events, and situations rather than abstract mental images, rules and roles..." (p. 558). On the basis of this and other related studies, the developmental aspects of students' cognitive abilities must be considered in curriculum development. However, student performance could also be explained in terms of the knowledge of relevant content. Certainly some attention must be paid to teaching specific science content in any proposal to change science education curricula.

REFERENCE

Piaget, J. The Psychology of Intelligence. Patterson, NJ: Littlefield, Adams and Co., 1966.

Schafer, L. E. and J. L. Byers. "The Effectiveness of Cue Fading in Teaching Kindergarten Children to Serial Order." Journal of Research in Science Teaching, 12(3):281-292, 1975.

Descriptors--*Cues; Educational Research; *Kindergarten Children; Learning; Learning Theories; *Learning Processes; *Primary Education; Science Education; *Serial Ordering

Expanded Abstract and Analysis Prepared Especially for I.S.E. by Willis J. Horak, University of Arizona.

Purpose

This study was designed to investigate the effects of instruction on the acquisition of serial ordering abilities of young children. Specifically, it dealt with the effectiveness of cuing and cue fading in an individualized situation upon the acquisition, retention, and transfer of the ability to insert objects into ordered sets. Additionally the study was viewed as an attempt to expand the theoretical basis for explaining the development of serial ordering abilities in young children. This was to be accomplished by ascertaining whether the ability to serial order is greatly influenced by attention factors such as learning to attend to relevant task characteristics.

Rationale

The rationale for this study is developed around two educational theories. The first one cited is based upon the developmental psychology of Piaget. Piaget is interpreted by many educators to contend that limited specific training cannot replace general types of instruction in fostering cognitive development. Many studies have, however, cast doubt on this contention. Consequently this study was designed to extend the tests of Piaget's contention by providing specific instruction in serial ordering.

The second theory is identified as the "American learning theory" by the researchers and defined and characterized as 1) placing emphasis on corrective feedback, 2) paying attention to relevant task stimuli,

3) including cuing and cue fading in instruction, and 4) sequencing instruction from least difficult to most difficult, thus 5) insuring a high incidence of success throughout the instruction. The research related to this approach is based upon the cited studies of Bloom and his colleagues and other studies by Gagne and his colleagues which are not explicitly cited.

Research Design and Procedures

The experimental design utilized in this study was the pretest-posttest control group design with repeated measures. For this study children were randomly assigned to treatment or control or special control conditions after they were determined to be on Piaget's second stage or third stage in the development of serial ordering capabilities. The second stage of serial ordering is characterized by children's ability to order objects by trial and error, and by their inability to insert a disarranged set of objects into an ordered set. The experimental group contained 15 children in the second stage of seriation capabilities and the control group contained 17 children in the second stage of seriation capabilities. All children were selected from a kindergarten class in a rural community.

In addition, children determined to be in Piaget's third stage were considered a special control group and given no instruction but received the identical post test 132 days after initial pretesting. These children were then compared to the original stage two experimental and control group children.

Instruction consisted of three 30-minute sessions with each child. All of the training was focused on the insertion capability. The first and the second sessions consisted of 45 tasks and 30 tasks respectively and were related to seriation of sticks. The third session consisted of 24 tasks and was related to seriation of cards upon which parallel lines had been drawn. All sessions utilized cuing and cue fading. The cues for the sticks dealt with the incremental sizes of the sticks and the cues for the parallel lines dealt

with the relationship of width of lines to the number of lines on each card.

Post-tests were administered one, eight, and 132 days after instruction to test retention. These tests consisted of 11 tasks classified as either near or far transfer tasks which were administered during the first and the second testing session and eight tasks classified as either far or far-far transfer which were administered during the third testing session. Children either had to serially order a set of materials or to insert objects into a previously ordered set of materials.

Findings

The collected data were analyzed utilizing a repeated measures multivariate analysis of variance. The analysis revealed significant (0.05 level) treatment, posttest (retention), and test type (transfer) main effects. Additionally it revealed significant treatment X test type and treatment X posttest X test type interactions. Univariate repeated measures analysis of variance techniques were utilized as post hoc procedures. On the near transfer data a significant treatment main effect was revealed with the experimental group's overall transfer mean significantly greater than the control group's overall transfer mean. On the far transfer data a significant posttest main effect was revealed. Additional analyses were conducted with the special control group which consisted of those children ascertained to be in stage three seriation before the treatments, and with the experimental and control groups separately. The results of these special analysis revealed that 132 days after instruction the experimental and special control groups did not differ significantly on any tasks, but that the control and the special control groups did differ significantly on the near and the far transfer tasks.

Interpretations

The instructional treatment utilizing cueing and cue fading produced substantial durable changes in children's abilities to perform specific serial ordering tasks. Therefore, apparently the acquisition of serial ordering capabilities depends in part on learning and not solely on the unfolding of some internal developmental structure or mechanism.

ABSTRACTOR'S ANALYSIS

In the introduction of this article the authors cite the work of Piaget and colleagues as influencing the rationale for the study. They contend that the results may indicate that specific limited training can supplant the role played in cognitive development by massive general types of experience. The analysis of the data reveals that this is apparently true. However, if the theoretical basi of the study lies in the cognitive developmental ideas of Piaget, it appears as though different treatments of the data or analysis might have been attempted.

Inhelder and Piaget (1964, p. 249) have indicated that serial operations are simply an interiorized result of previous activities. Their origin must be sought in sensori-motor schemata rather than in a purely perceptual schema. The operational schema of seriation is anticipatory. Students realize in advance that, when they are faced with an ordering task or an insertion task, that, by choosing the smallest element that remains in a set, they will eventually build a series in which each term is larger than the preceding ones. With this operational view of seriation there is no reason to assume that within the anticipatory schema scoring 80 percent on a group of seriation tasks is any better than scoring 66 percent on those same tasks. What is important, as far as cognitive development, is the ascertained stage of development or the change in the stage of development. Similarly, even though significantly different than another group's score, does a score of 51.9 percent on far-far transfer tasks imply superiority in cognitive developmental levels? Maybe or maybe not. If we are

interested in cognitive development we must, as Inhelder, Sinclair, and Bovet (1974, p. 7) state, "realize that it is necessary to make a distinction between what the subject learns from the point of view of the form or the logical-mathematical framework of the concepts—the reflective abstraction—and what he learns from the point of view of the content of the concepts—the simple abstraction." If we are concerned about the former, then we must be concerned with the level of development of the child which most times is more than percent of correct solutions. If the content of the concepts is important then the percent of correct responses is an appropriate concern.

This study seemed to not address either specific issue. It initially assigned students to groups based upon their stages in the development of serial ordering and then did not identify the stage of development of the children after the instructional treatment. Additionally, the percentage scored on the tasks identified as far transfer and far-far transfer reveal percentages that would lead one to doubt whether children initially identified as stage three were actually in stage three. Or, if the children are truly in stage three, then perhaps the specific items are requiring more than seriation ability to complete. Similarly, if we are interested in the content of the concept then initially grouping the children by levels or stages of cognitive development on seriation tasks limits the generalizability of our findings. If the cuing and cue fading instructional treatment does affect seriation ability, it may be meaningful to know if it has parallel or disordinal interaction effect on children at different cognitive levels.

The study is more meaningful when related to other studies in the area of concept learning rather than cognitive development. Much of the research in this area has addressed itself to the dual problems of transfer and retention. This study expands that base of knowledge by utilizing a cuing and cue fading instructional program along with seriation activities. Additionally the identification of the levels of transfer of the post test tasks is most useful. The complete

description of how the sample was obtained and how the data were analyzed is most useful. It is, however, impossible to tell from the article what was the content of part one and part two of the post test. This may be important not only from a development point of view but also from a standpoint of analyzing cuing and cue fading as a viable instructional method. We must in this case be concerned not only with all levels of development but also with the specific types of post test items.

From a retention point of view, this study is most beneficial. Many previous retention studies have been conducted from a memory standpoint with retention being viewed as influenced by the processes of consolidation of learned material by rehearsal and by activation of a type of memory trace (Herriot, Green, and McConkey). They believe that a memory trace undergoes deterioration over a period of time, thus leading to forgetting especially if the elapsed time involved activity. The long-range retention results of this study appear to indicate that actually not much forgetting did occur. However, a perusal of the slopes of the lines representing the scores for the near and far transfer seem to indicate that eventually the control and the experimental groups scores will converge.

In corclusion, the fact that two research bases for this study were applied makes it hard to ascertain the full importance of the study for either area of research. Many procedures that are most useful in one area are not applicable to the other area. Due to the fact that percentages of correct responses were the reported data, I would question the statement of the researchers that, contrary to Piaget's notion, specific short periods of instruction lead to expanded cognitive development. In his article "Cognitive Development and the Learning of Elementary Concepts," J. F. Wohlwill aptly sums up the cited research dilemma when he stated, "It is not only the acquisition of a new response, such that a child might learn about something that he did not know before. It is not the acquisition of facts or knowledge generally, but it is a matter of giving up one type of response which is extinguished while another one is being developed. I think

that when a learning interpretation is advanced, this has to be borne in mind."

REFERENCES

Herriot, P.; J. M. Green; and R. McConkey. Organization and Memory: A Review and a Project in Subnormality. London: Metuen and Co. Ltd., 1973.

Inhelder, B. and J. Piaget. The Early Growth of Logic in the Child. New York: W. W. Norton and Company, Inc., 1964.

Inhelder, B.; H. Sinclair; and M. Bovet. Learning and the Development of Cognition. Cambridge, MA: Harvard University Press, 1974.

Wohlwill, J. F. "Cognitive Development and the Learning of Elementary Concepts." Journal of Research in Science Teaching, 2:222-226, 1964.

TEACHER EDUCATION

53/5457

Chrouser, William H. "Outdoor vs. Indoor Laboratory Techniques in Teaching Biology to Prospective Elementary Teachers." Journal of Research in Science Teaching, 12(1):41-48, 1975.

Descriptors--*Biology; Educational Research; *Elementary School Teachers; Higher Education; *Instruction; *Laboratory Techniques; Preservice Education; Science Education; *Teacher Education

Expanded abstract and analysis prepared especially for I.S.E. by William R. Brown, Old Dominion University.

Purpose

This empirical study was conducted in order to justify the value of outdoor experiences in the education of prospective elementary teachers. Four null hypotheses were tested: there is no significant difference in mean scores earned by students in an indoor group and students in an outdoor group: (1) on the understanding of the social aspects of science, (2) in the understanding of selected biological principles, (3) on understanding of science as process, and (4) in critical thinking.

Rationale

The investigator stated that descriptive surveys and historical studies tended to provide the research base for outdoor education. Several studies were cited that formed the literature base for this investigation. The rationale to investigate the effects of outdoor education was primarily based on the writings of Clifford Knapp (1972). He argued that pedalogical methods and the nature of the learner must be considered in research dealing with outdoor education.

Research Design and Procedure

The variate was method of instruction. Each class was divided into an outdoor group and an indoor group. Differences in instruction centered on types of activities with all other "methods" being the same. The indoor group performed only those exercises with no outdoor activity while the outdoor group performed activities which were in the outdoors. All activities dealt with the same principles.

The criterion variable was performance. This was assessed by gain from pre to post test. Tests measured social aspects of science, understanding of biological principles, science as process, and critical thinking.

The subjects (n = 40) for the investigation were prospective elementary teachers enrolled in a university biology course. The study occurred during the spring and summer quarters. The treatment was of approximately ten weeks duration. The design model was pretest-posttest for both experimental and control groups with outdoor activities as the treatment. This may be represented as

$$\begin{array}{c} 0 __ 0 \\ 0 \ X \ 0 \end{array}$$

Instruments used were the Test on the Social Aspects of Science (TSAS). A KR-20 reliability estimate of 0.71 was reported as established by a prior investigator.

The Methods and Procedures of Science: An Examination (MPSE) emphasized the process of science (especially biology), logic used in scientific manipulation, and understanding of terms. A reliability index of 0.802 was cited.

A reliability of 0.85-0.93 was reported for the Watson-Glaser Critical Thinking Appraisal (WGCTA). The five subtests are inference, recognition of assumption, deduction, interpretation, and evaluation of arguments.

The fourth instrument used was the Understanding of Selected Biological Principles: An Examination (USBPE). A KR-20 coefficient of 0.904 was calculated. Validity was established by matching test-items to principles taught.

All four instruments were administered as pretests and as posttests.

Analysis of covariance (ANCOVA) was used for results of the USBPE, the MPSE, and the WGCTA. Data from the TSAS were analyzed by Chi-Square Test for Independence. The TSAS was analyzed using all 40 subjects. The data from the other tests were analyzed using spring (n = 28) and

summer (n = 12) separately. This was done due to differences in class size and because of different levels of students (spring = sophomores through seniors; summer = seniors and graduate students).

Findings

Hypotheses are fully stated in the first section of this abstract.

Hypothesis 1 - Social Aspects; TSAS; Chi-Square; n = 40. Significant differences (.05) were found between the indoor group and the outdoor group on eight of the 19 items dealing with science and technology as they interact with each other and with society. The part concerning the nature of the scientific enterprise showed a significant difference on 14 of the 21 items. There were significant differences on 7 of the 12 items relating to the social responsibilities of science and scientists. Since significant differences were found on more than half the items, the null hypothesis was rejected.

For hypotheses 2-4, data were treated separately for spring (n = 28) and summer (n = 12).

Hypothesis 2 - Principles; USBPE; ANCOVA. No significant differences between adjusted means were found. Hypothesis 2 could not be rejected.

Hypothesis 3 - Process; MPSE; ANCOVA. This hypothesis was rejected.

Hypothesis 4 - Critical Thinking; WGCTA; ANCOVA. This hypothesis could not be rejected.

Interpretations

The investigator concluded that an outdoor laboratory emphasis is more effective than indoor emphasis in helping students achieve understanding of: (1) the social aspects of science, (2) selected biological principles, and (3) science as process. Understanding of biological principles in general and critical thinking ability did not differ as related to indoor-outdoor emphasis.

The interpretation of biological principles was reported as selected and as general based on the analysis of five short recall quizzes representing selected principles and the USBPE as a whole representing general principles.

Several subjective behaviors were reported. Based on empirical data and subjective observations, the investigator implied that there is no substitute for a field laboratory. He also stated that preservice and inservice teachers "need" to be trained in depth by specialists in content subject matter. This training should be aided by the use of the outdoors.

ABSTRACTOR'S ANALYSIS

The comments in this section are in three categories. These are comments on the (1) design of the investigation, (2) written report, and (3) assessment of the interaction between curriculum and instruction.

Design of the Investigation

The total n was 40 students, 28 spring quarter and 12 summer quarter. This appears to be an insufficient number of subjects to use to draw conclusions. It is also assumed by this reviewer that the same instructor, who was also the investigator, taught both the indoor and outdoor groups. Interaction variables related to the instructor/investigator could influence results. Perhaps the use of another instructor and additional subjects could lend support to the generalizations reported.

A second design problem was the lack of reported randomization. How was each class divided into an outdoor and an indoor group? The pretest-posttest, control-experimental group design is greatly strengthened by random assignment of subjects to each group.

It is assumed by reading the report that each treatment lasted approximately ten weeks. Is it reasonable to expect gain in such complex areas as understanding the social aspects of science, understanding science as process, and in the ability to think critically in two and a half months?

Prior experiences in science would most likely affect results. Was this the only science course for the subjects? No measures were reported to control or account for learner variables, teacher variables, or instructional variables. The environmental variable (indoor vs. outdoor) could very well be influenced by other classes of variables (Baker and Schultz, 1972, pp. 46-49).

The use of Chi-Square and ANCOVA were appropriate for the types of variables identified (Baker and Schultz, 1972, p. 180). A question not addressed in the report was why was hypothesis one rejected based on more than half the items of the TSAS showing significant differences? Perhaps the TSAS should have been used to assess three separate sub-sections and three sub-hypotheses, especially since data were reported in three separate tables.

Written Report

The specification of number of subjects was not made until the analysis of data section. It is helpful to a reader to see this critical information under the design and procedures section.

Were students assigned to group by random methods? If so, this should be stated clearly in the design section.

How long was the treatment? At first reading this reviewer was under the impression that the treatment was approximately 20 weeks instead of 10.

Validity and reliability estimates help the reader to make judgments as to the appropriateness of instruments. Comments on validity were cited for the WGCTA and the USBPE only. Reliability estimates were reported for all tests, but the WGCTA values of 0.85-0.93 and the MPSE values of 0.802 were not linked to a specific index such as KR-20. Interpretation of reliability cannot be made and comparisons made without knowing the index used.

In the reporting and analyzing of data, results of the TSAS were adequately described. No data were reported for the other three instruments. Acceptance or rejection of three hypotheses were determined, but no data were presented that would enable a reader to "check" calculations.

In interpreting the data from the USBPE, the investigator separated selected principles as measured by five quizzes with items from the USBPE, from general principles as assessed by the intact USBPE. The investigator stated that "it may be that short recall test results on laboratory quizzes differ significantly from longer recall tests results on the total USBPE instrument." Can it be said with confidence that specific vs. general principles were being assessed or were time and the amount of content the variables? If quizzes were to be used in assessment and interpretation, then hypotheses should have been stated to accommodate these instruments.

Several subjective behaviors were reported in conclusions. Since these factors were not part of the investigation as detailed, should they be reported? The inclusion of subjective behaviors is contrary to the investigator's rationale for the need for empirical research.

Interaction Between Curriculum and Instruction

It appears to this reviewer that research in the areas of instruction can be strengthened by a careful consideration of what instruction actually is. Three definitions may aid in this consideration (White, 1977).

Curriculum-A plan, a structured series of intended student outcomes or a set of planned experiences indicating that which has been identified as worthwhile for students to learn.

Program-A specific area of instruction (e.g., English, art, biology).

Instruction-The behavior of teachers and the strategies selected to promote the objectives of the curriculum or program.

Curriculum and program evaluation should be accomplished through the consistent and logical use of statements previously established in curriculum decision making (White, 1977). For example, if a biology program decision is to "develop critical thinkers," then the program or plan should be reviewed to see that experiences are included to do what has been specified. How well students perform on a test of critical thinking should not necessarily influence modification of the plan. If development of critical

thinkers is specified as a curriculum/program objective, then teachers must select strategies and display behaviors to develop this goal. If assessment of students indicates critical thinkers are not developing, then instruction should be the first consideration. Instruction, not curriculum, is best evaluated by student performance.

In this investigation, Chrouser examined the environmental variable of laboratory setting. The two settings were indoor and outdoor. He found no significant change in critical thinking or in the understanding of selected biological principles. Assuming that these two areas are stated as curriculum/program goals, the immediate area of concern is instruction. What can be done to help students become critical thinkers and to "understand" selected biological principles? What other instructional variables such as texts, classroom organization, modes of instruction, content organization, learner practice, grouping of students, frequency of assessment, types of student responses and feedback enter into student outcomes (Baker and Schultz, 1973, pp. 46-49)? Answers to these complex instructional questions can only be developed by examining teacher behaviors and strategies. Chrouser's study definitely contributes a "bit" of information to assessment of the effectiveness of the learning environment.

REFERENCES

Baker, Robert and Richard Schultz (Eds.). Instructional Product Research, Cincinnati, Ohio: Van Nostrand Co., 1972.

Knapp, Clifford. "Some Challenges in Outdoor Education." Perspectives on Outdoor Education: Readings. (Edited by George Donaldson and Oswald Goering.) Dubuque, Iowa: William C. Brown Co., 1972, Chap. 4, 116-121.

White, Edwin. "Principles of Curriculum Development." Handbook for Elementary Science Education. (Edited by William R. Brown.) Norfolk, Virginia: Science Education Center of Old Dominion University, 1977.

Greenwood, Gordon E. and Richard R. Renner. "Student Ratings of College Teaching: Their Validity and Use in Administrative Decision Making." Science Education, 59(4):493-498, 1975.

Descriptors--College Faculty; Educational Research; *Faculty Evaluation; Higher Education; *Literature Reviews; Science Education; *Student Opinion; *Teaching Quality; *Validity

Expanded Abstract and Analysis Prepared Especially for I.S.E. by Herbert A. Smith, Colorado State University.

Purpose

The purpose appeared to be to conduct a survey of the literature with respect to student ratings of college teaching with some assessment of the rating's validity and its potential for use in administrative decision making.

Rationale

Student evaluations of faculty have become an increasingly important component of overall faculty evaluations. The validity of this procedure is certainly an important educational question, particularly in higher education where tenure and promotion through the professorial ranks has critical career implications. The report by Greenwood and Renner is an attempt to assess the progress which research has made in addressing the validity question of student ratings of faculty performance and the implications which may exist for administrative decision making.

Research Design and Procedure

The paper can hardly be described as a report of research; it is rather a report on a search of the literature with respect to the validity of student ratings of college teaching. It appears to be a relatively comprehensive search for investigations bearing on this relatively limited, but important, topic.

Findings and Conclusions

Perhaps the best feature of the paper relates to the questions which it raises in its commentaries on some of the research investigations on which it reports. These questions could be phrased as follows; however, the language is that of the reviewer and not of the authors:

1. Would the findings of studies on the validity of student evaluations be independent of the discipline in which courses are taught, graduate or undergraduate levels of instruction, and different kinds of institutions of higher education?

2. What is an appropriate criterion measure to employ in determining the validity of student evaluations?

3. Should student evaluation instruments reflect the character and climate of the learning atmosphere and process, while other instruments are used to measure the achievement of course objectives?

The major conclusion of the study was that, in general, the studies addressing the question of the validity of student evaluations to date are inconclusive. There is a further implication that the contributions which they can make to an overall administrative determination of faculty effectiveness is quite limited. They inject a rather strong cautionary note on the use of such evaluations with respect to the determination of faculty tenure, promotion and salary increments. They advocate the use of substantial other inputs.

ABSTRACTOR'S ANALYSIS

The paper clearly addresses itself to a very important issue. Nevertheless, it tends to leave the reader unsatisfied and with the feeling that the analysis was too shallow and clearly did not exhaust the subject. The report strikes one as being a series of provocative and informative quotes interspersed with commentary.

The question which they raise with respect to an appropriate validity criterion is especially critical. As Greenwood and Renner point out, the validity of student evaluations of faculty performance will be demonstrated when and if it can be shown that there is a high correlation between student evaluations and other more objective measures of gains in learning which have resulted. They further point out that very few studies exploring this relationship have been reported. They also indicate that most studies of student ratings of instruction relate the ratings to achievement as measured by grade earned at the end of a course.

There appears to be a very serious defect in using grades as a criterion for validity determination of student evaluations but the authors do not pursue the matter. For example, one would anticipate as tenable an hypothesis that students' evaluations of instructors are highly correlated with the grades, or the anticipated grades they are likely to receive at the end of the courses. If this is true, the whole process of faculty evaluation by students would tend to be invalidated when grades are used as an external criterion. It is a reasonable speculation that the whole problem of grade inflation which has been demonstrated to be widespread in recent years is highly related to, or in fact caused by, this particular phenomenon. If faculty performance is to be judged by student evaluations and if student evaluations are determined to a large extent by the grades they receive, it is obviously in the instructor's best interest to award high grades.

Unfortunately neither high grades nor high faculty evaluations are necessarily a guarantee of high level achievement. There is some evidence presented in the article which supports this position. The authors report a study in which a rather astonishing -.75 correlation between students' ratings and students' achievement in calculus was found. A similar study based on evaluation of French instructors yielded a -.50 correlation. Such findings raise very serious questions about the long-term influence of any widespread use of student evaluations in ways which can have a serious impact on the professional careers of faculty members. The substance of the question is, "Will faculty in higher

education be employed for their affability, congeniality and generosity in marking or for what their students are to learn?" Unfortunately, the report by Greenwood and Renner does not delve deeply into any such questions as the one described above, or into many others which could be natural outgrowths of their survey.